Volume 101, Number 3, May–June 1996
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. **101**, 243 (1996)]

The Nucleic Acid Database: Present and Future

Volume 101 Number 3 May–June 1996

Helen M. Berman, Anke Gelbin, Lester Clowney,

In addition to coordinate data, information relevant to the crystallographic experiment is abstracted from the primary literature for inclusion into the database. These include crystallization conditions, refinement statistics and data collection statistics. Other derived information, such as the distances, angles, torsion angles, and base morphology parameters, is calculated from the coordinate data and placed in the database. Tables 2a and 2b list summaries of the information currently in the NDB.

Table 1. NDB holdings as of October 1995
408 structures (390 released)

Structure Type	Number
A-DNA	51
DNA/RNA Hybrid	11
A-RNA	10
DNA-Drug Complexes	93
B-DNA	66
RNA-Drug Complexes	19
Z-DNA	47
t-RNA	10
Unusual DNA	21
Protein-Nucleic Acid Complexes	66
Unusual RNA	14

Table 2a. Primary experimental information stored in the NDB

Structure summary[a]	Descriptor NDB, PDB, and CSD names Coordinates available (yes/no) Modifiers (yes/no) Mismatches (yes/no) Drugs (yes/no)
Structural description[a]	Sequence Structure type (A/B/Z/RH/U/P) Description of modifiers of base, phosphate, and sugar Description of base mismatch Name and binding type of drug Description of base pairing Description of contents of asymmetric unit
Citation[a]	Authors Title Journal Volume Pages Year
Crystal data[a]	Cell dimensions Space group
Data collection description[a]	Source of radiation Data collection device Radiation wavelength Temperature Resolution range Total and unique number of reflections
Crystallization description[a]	Method Temperature pH value Composition of solutions
Refinement information[a]	Method Program Number of reflections used for refinement Data cutoff Resolution range R-factor Refinement of temperature factors and occupancies
Coordinate information[b]	Atomic coordinates, occupancies and temperature factors for asymmetric unit Coordinates for symmetry related strands Symmetry related coordinates in unit cell (packing) Orthogonal or fractional coordinates

Table 2b. Derivative information stored in the NDB

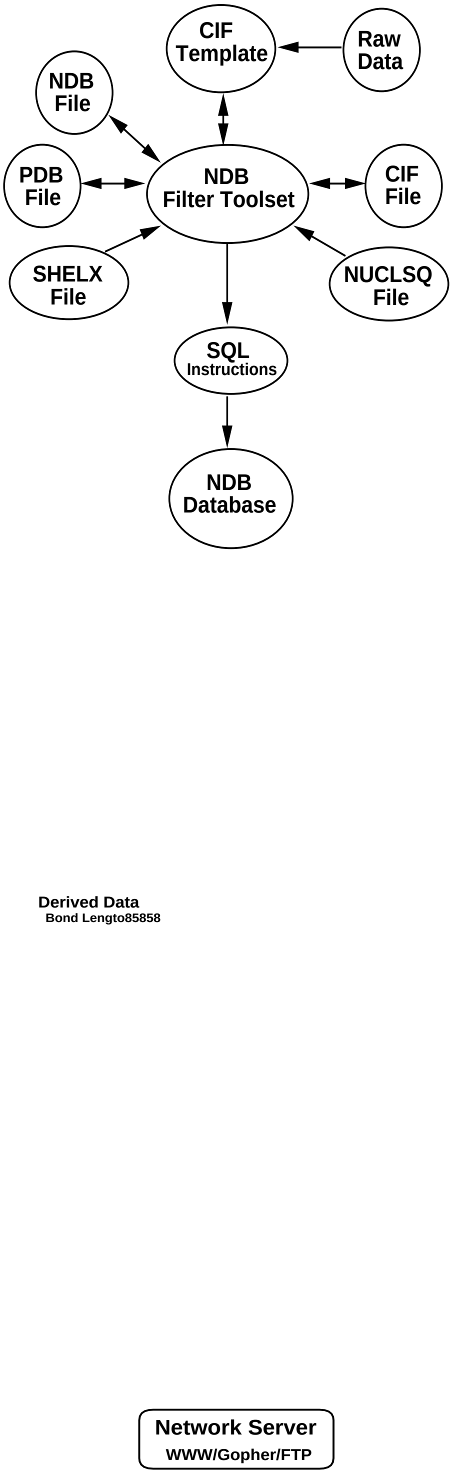
Distances[a]	Chemical bond lengths Virtual bonds (involving phosphorus atoms)
Torsions[c]	Backbone and side chain torsion angles Pseudorotational parameters
Angles[a]	Valence bond angles Virtual angles (involving phosphorus atoms)
Base morphology[a]	Parameters calculated by different algorithms

[a] Reports can be generated in either ASCII or LATEX.
[b] Reports can be generated as an NDB or PDB coordinate file, a Kinemage template, or as PostScript molecular graphics.
[c] Parameters can be displayed in both LATEX or ASCII tables, or as a PostScript conformation wheel.

3. Data Processing

3.1 Data Entry and Integrity Checks

The scheme for data processing is given in Figs. 1a and 1b. A set of filter programs have been developed that allow this process of data entry and integrity checking to be highly automated. A key feature of the system is the use of a template based on mmCIF. A template is a CIF data file that includes definition and example


CIF
Template
Raw
Data
NDB
File
PDB
File
NDB
Filter Toolset
CIF
File
SHELX
File
NUCLSQ
File
SQL
Instructions
NDB
Database
Derived Data
Bond Lengto85858
Network Server
WWW/Gopher/FTP

which have been determined by a particular author. Two examples of the use of structure selection constraints are presented in Tables 3a and 3b.

It is possible to use either the menu driven interface to NDBquery or the WWW forms based system to generate selection constraints. The advantage of the latter method is that it places no restrictions on the user other than the ability to use the World Wide Web using either Netscape or Mosaic. A sample query using the WWW access is shown in Fig. 2.

4.2 Report Generation

Once the selection constraints are defined, a large variety of reports can be generated that describe any of the properties that are stored in the database. The simplest type of report is the list of coordinates for the selected structures. In addition, the NDBquery program produces reports in a wide variety of formats. Tabular reports such as those shown in Fig. 3 can be produced in either ASCII or PostScript formats.

Graphical reports relating any two properties can be generated. It is possible to produce scatter charts, histograms, and pie charts that can be used to analyze the properties of the structures contained within the database. These report features were used to examine the frequency distributions as well as the correlations of torsion angles of the three classes of DNA duplexes. In order to automate this type of survey, batch query capabilities were built into the system. Examples of graphical outputs are shown in Fig. 4.

The NDBquery program also produces molecular graphics in a variety of formats. Structures can be depicted using color codes for the properties of the atoms or residues. Automatic packing pictures are generated in PostScript format using NDBquery and in raster form using NDBview [3]. Various types of representations, including ball and stick and Van der Waals spheres, are available (Fig. 5).

There are provisions for detailed formatting so that a complete set of publication quality reports for a set of structures can be produced. To simplify the query process, some standard and commonly used queries are saved and made available for the user. In addition, the user may save her own queries to be used repeatedly for a particular project.

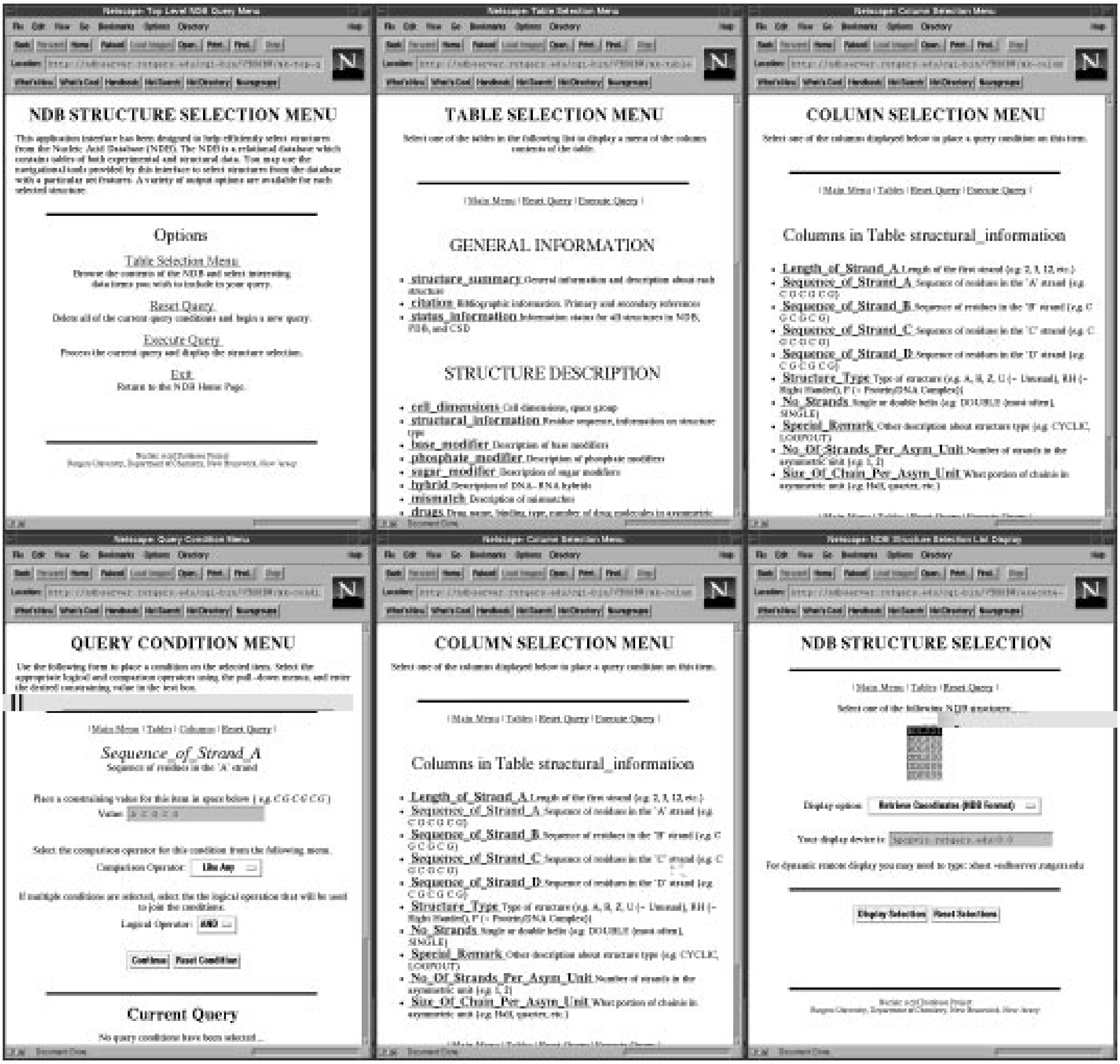
The WWW forms based interface also allows for report generation. Coordinates may be retrieved in mmCIF, NDB or PDB format. It is also possible to retrieve an Atlas page (see later) and to view the structure using a dynamic viewer. The latest version of the WWW Interface can also create tabular reports based on any of the features contained in the database.

Table 3a. Example 1: Structure selection of B-DNAs containing the residue sequence ‘‘C G C G ’’ without base modifiers, mismatches, or drugs

Table	Property	Operator	Operand	Logical
structural_information	structure_type	=	B	AND
structural_information	Sequence_of_Strand_A	like	%CGCG%	AND
structure_summary	base_modifier	is null		AND
structure_summary	mismatch	is null		AND
structure_summary	drug	is null		AND

Table 3b. Example 2: Structure selection of B-DNAs with resolution ≤1.9 Å and R factors <0.17 by authors A. Rich, R. E. Dickerson, or O. Kennard

Table	Property	Operator	Operand	Logical
structural_information	structure_type	=	B	AND
r_factor	Up_Lim_Resol_Ref	≤	1.9	AND
r_factor	R_Value	<	0.17	AND
citation	authors	like	R. E. Dickerson	OR
structural_information	structure_type	=	B	AND
r_factor	Up_Lim_Resol_Ref	≤	1.9	AND
r_factor	R_Value	<	0.17	AND
citation	authors	like	A. Rich	OR
structural_information	structure_type	=	B	AND
r_factor	Up_Lim_Resol_Ref	≤	1.9	AND
r_factor	R_Value	<	0.17	AND
citation	authors	like	O. Kennard	


Netscape: Top Level NDB Query Menu
Netscape: Table Selection Menu
Netscape: Column Selection Menu
Netscape: Query Condition Menu
Netscape: Column Selection Menu
Netscape: NDB Structure Selection List Display

NDB STRUCTURE SELECTION MENU

This application interface has been designed to help efficiently select structures from the Nucleic Acid Database (NDB). The NDB is a relational database which contains tables of both experimental and structural data. You may use the navigational tools provided by this interface to select structures from the database with a particular set of features. A variety of output options are available for each selected structure.

Options

Table Selection Menu
Browse the contents of the NDB and select interesting data items you wish to include in your query.

Reset Query
Delete all of the current query conditions and begin a new query.

Execute Query
Process the current query and display the structure selection.

Exit
Return to the NDB Home Page.

Nucleic Acid Database Project
Rutgers University, Department of Chemistry, New Brunswick, New Jersey

TABLE SELECTION MENU

Select one of the tables in the following list to display a menu of the column contents of the table.

| Main Menu | Reset Query | Execute Query |

GENERAL INFORMATION

- structure_summary General information and description about each structure
- citation Bibliographic information. Primary and secondary references
- status_information Information status for all structures in NDB, PDB, and CSD

STRUCTURE DESCRIPTION

- cell_dimensions Cell dimensions, space group
- structural_information Residue sequence, information on structure type
- base_modifier Description of base modifiers
- phosphate_modifier Description of phosphate modifiers
- sugar_modifier Description of sugar modifiers
- hybrid Description of DNA–RNA hybrids
- mismatch Description of mismatches
- drugs Drug name, binding type, number of drug molecules in asymmetric

COLUMN SELECTION MENU

Select one of the columns displayed below to place a query condition on this form.

| Main Menu | Tables | Reset Query | Execute Query |

Columns in Table structural_information

- Length_of_Strand_A Length of the first strand (e.g. 2, 3, 12, etc.)
- Sequence_of_Strand_A Sequence of residues in the 'A' strand (e.g. C G C G C G)
- Sequence_of_Strand_B Sequence of residues in the 'B' strand (e.g. C G C G C G)
- Sequence_of_Strand_C Sequence of residues in the 'C' strand (e.g. C G C G C G)
- Sequence_of_Strand_D Sequence of residues in the 'D' strand (e.g. C G C G C G)
- Structure_Type Type of structure (e.g. A, B, Z, U (= Unusual), RH (= Right Handed), P (= Protein/DNA Complex))
- No_Strands Single or double helix (e.g. DOUBLE (most often), SINGLE)
- Special_Remark Other description about structure type (e.g. CYCLIC, LOOPOUT)
- No_Of_Strands_Per_Asym_Unit Number of strands in the asymmetric unit (e.g. 1, 2)
- Size_Of_Chain_Per_Asym_Unit What portion of chain is in asymmetric unit (e.g. Half, quarter, etc.)

| Main Menu | Tables | Reset Query | Execute Query |

QUERY CONDITION MENU

Use the following form to place a condition on the selected item. Select the appropriate logical and comparison operators using the pull-down menus, and enter the desired constraining value in the text box.

| Main Menu | Tables | Columns | Reset Query |

Sequence_of_Strand_A
Sequence of residues in the 'A' strand

Place a constraining value for this item in space below (e.g. *C G C G C G*)
Value: A C G C G

Select the comparison operator for this condition from the following menu.
Comparison Operator: Like Any

If multiple conditions are selected, select the the logical operation that will be used to join the conditions.
Logical Operator: AND

Continue | Reset Condition

Current Query

No query conditions have been selected ...

COLUMN SELECTION MENU

Select one of the columns displayed below to place a query condition on this form.

| Main Menu | Tables | Reset Query | Execute Query |

Columns in Table structural_information

- Length_of_Strand_A Length of the first strand (e.g. 2, 3, 12, etc.)
- Sequence_of_Strand_A Sequence of residues in the 'A' strand (e.g. C G C G C G)
- Sequence_of_Strand_B Sequence of residues in the 'B' strand (e.g. C G C G C G)
- Sequence_of_Strand_C Sequence of residues in the 'C' strand (e.g. C G C G C G)
- Sequence_of_Strand_D Sequence of residues in the 'D' strand (e.g. C G C G C G)
- Structure_Type Type of structure (e.g. A, B, Z, U (= Unusual), RH (= Right Handed), P (= Protein/DNA Complex))
- No_Strands Single or double helix (e.g. DOUBLE (most often), SINGLE)
- Special_Remark Other description about structure type (e.g. CYCLIC, LOOPOUT)
- No_Of_Strands_Per_Asym_Unit Number of strands in the asymmetric unit (e.g. 1, 2)
- Size_Of_Chain_Per_Asym_Unit What portion of chain is in asymmetric unit (e.g. Half, quarter, etc.)

| Main Menu | Tables | Reset Query | Execute Query |

NDB STRUCTURE SELECTION

| Main Menu | Tables | Reset Query |

Select one of the following NDB structures:

Display option: Retrieve Coordinates (NDB Format)

Your display device is: [illegible]

For dynamic remote display you may need to type: xhost +ndbserver.rutgers.edu

Display Selection | Reset Selections

Nucleic Acid Database Project
Rutgers University, Department of Chemistry, New Brunswick, New Jersey

Fig. 2. Sequence for a simple query, i.e., choosing structures that contain the specific sequence ACGCG using the WWW Interface, version 2.0 (October 1995).

Beginning from the upper left:

a. The **Table Selection Menu** from the **NDB Structure Selection Menu** is chosen**.**

b. The **Structural_information** menu is selected from the **Table Selection Menu.**

c. **Sequence_of_Strand_A** is selected from the **Column Selection Menu**.

d. The desired sequence, A C G C G, is entered in capital letters with spaces separating each residue in the provided field. To move to the next step, the **Continue** bar, is selected.

e. Once all of the desired constraints are selected, **Execute Query** is pressed from the top of the Column Selection Menu.

f. A list of the NDB identifiers of the structures containing the sequence ACGCG is presented. The user may now:

Retrieve coordinates in NDB Format

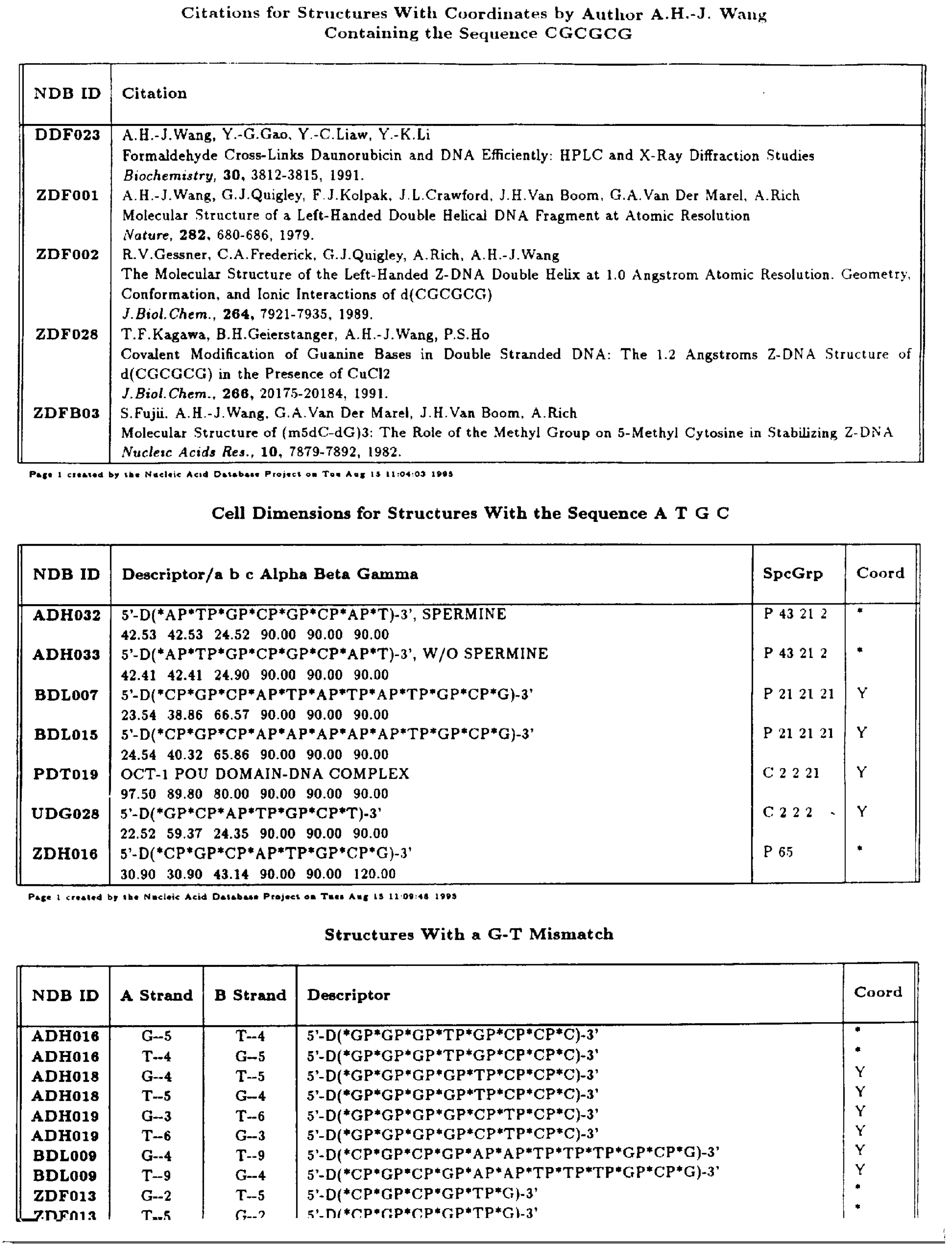
Retrieve coordinates and the bibliographic information in NDB Format (Full Entry)

Retrieve coordinates in PDB Format

Display the structure using a remote viewer (launching RasMol viewer on ndbserver)

Display the structure using a local viewer (launching your own viewer)

Display the Atlas Entry for the structure



Citations for Structures With Coordinates by Author A.H.-J. Wang Containing the Sequence CGCGCG

NDB ID	Citation
DDF023	A.H.-J.Wang, Y.-G.Gao, Y.-C.Liaw, Y.-K.Li Formaldehyde Cross-Links Daunorubicin and DNA Efficiently: HPLC and X-Ray Diffraction Studies *Biochemistry*, **30**, 3812-3815, 1991.
ZDF001	A.H.-J.Wang, G.J.Quigley, F.J.Kolpak, J.L.Crawford, J.H.Van Boom, G.A.Van Der Marel, A.Rich Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution *Nature*, **282**, 680-686, 1979.
ZDF002	R.V.Gessner, C.A.Frederick, G.J.Quigley, A.Rich, A.H.-J.Wang The Molecular Structure of the Left-Handed Z-DNA Double Helix at 1.0 Angstrom Atomic Resolution. Geometry, Conformation, and Ionic Interactions of d(CGCGCG) *J.Biol.Chem.*, **264**, 7921-7935, 1989.
ZDF028	T.F.Kagawa, B.H.Geierstanger, A.H.-J.Wang, P.S.Ho Covalent Modification of Guanine Bases in Double Stranded DNA: The 1.2 Angstroms Z-DNA Structure of d(CGCGCG) in the Presence of CuCl2 *J.Biol.Chem.*, **266**, 20175-20184, 1991.
ZDFB03	S.Fujii. A.H.-J.Wang, G.A.Van Der Marel, J.H.Van Boom, A.Rich Molecular Structure of (m5dC-dG)3: The Role of the Methyl Group on 5-Methyl Cytosine in Stabilizing Z-DNA *Nucleic Acids Res.*, **10**, 7879-7892, 1982.

Page 1 created by the Nucleic Acid Database Project on Tue Aug 15 11:04:03 1995

Cell Dimensions for Structures With the Sequence A T G C

NDB ID	Descriptor/a b c Alpha Beta Gamma	SpcGrp	Coord
ADH032	5'-D(*AP*TP*GP*CP*GP*CP*AP*T)-3', SPERMINE 42.53 42.53 24.52 90.00 90.00 90.00	P 43 21 2	*
ADH033	5'-D(*AP*TP*GP*CP*GP*CP*AP*T)-3', W/O SPERMINE 42.41 42.41 24.90 90.00 90.00 90.00	P 43 21 2	*
BDL007	5'-D(*CP*GP*CP*AP*TP*AP*TP*AP*TP*GP*CP*G)-3' 23.54 38.86 66.57 90.00 90.00 90.00	P 21 21 21	Y
BDL015	5'-D(*CP*GP*CP*AP*AP*AP*AP*AP*TP*GP*CP*G)-3' 24.54 40.32 65.86 90.00 90.00 90.00	P 21 21 21	Y
PDT019	OCT-1 POU DOMAIN-DNA COMPLEX 97.50 89.80 80.00 90.00 90.00 90.00	C 2 2 21	Y
UDG028	5'-D(*GP*CP*AP*TP*GP*CP*T)-3' 22.52 59.37 24.35 90.00 90.00 90.00	C 2 2 2	Y
ZDH016	5'-D(*CP*GP*CP*AP*TP*GP*CP*G)-3' 30.90 30.90 43.14 90.00 90.00 120.00	P 65	*

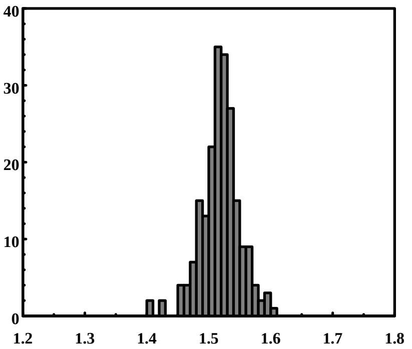
Page 1 created by the Nucleic Acid Database Project on Tues Aug 15 11:09:48 1995

Structures With a G-T Mismatch

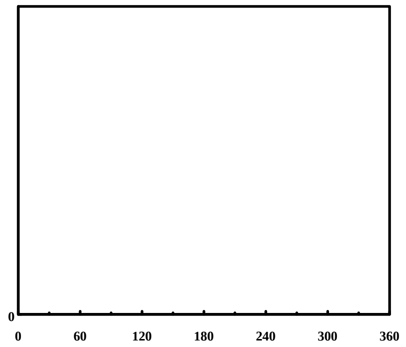
NDB ID	A Strand	B Strand	Descriptor	Coord
ADH016	G--5	T--4	5'-D(*GP*GP*GP*TP*GP*CP*CP*C)-3'	*
ADH016	T--4	G--5	5'-D(*GP*GP*GP*TP*GP*CP*CP*C)-3'	*
ADH018	G--4	T--5	5'-D(*GP*GP*GP*GP*TP*CP*CP*C)-3'	Y
ADH018	T--5	G--4	5'-D(*GP*GP*GP*GP*TP*CP*CP*C)-3'	Y
ADH019	G--3	T--6	5'-D(*GP*GP*GP*GP*CP*TP*CP*C)-3'	Y
ADH019	T--6	G--3	5'-D(*GP*GP*GP*GP*CP*TP*CP*C)-3'	Y
BDL009	G--4	T--9	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*TP*GP*CP*G)-3'	Y
BDL009	T--9	G--4	5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*TP*GP*CP*G)-3'	Y
ZDF013	G--2	T--5	5'-D(*CP*GP*CP*GP*TP*G)-3'	*
ZDF013	T--5	G--2	5'-D(*CP*GP*CP*GP*TP*G)-3'	*

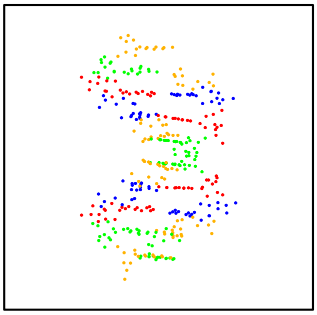
Fig. 3. Examples of Postscript reports created by NDBquery.

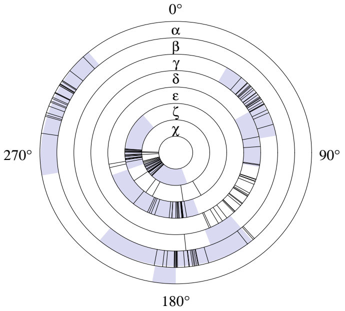
Z-DNA



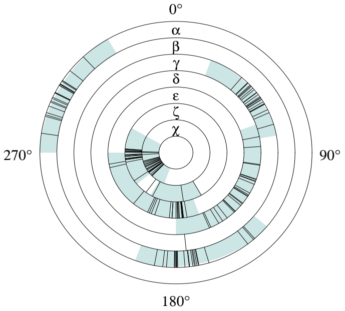
C5' - C4' distance



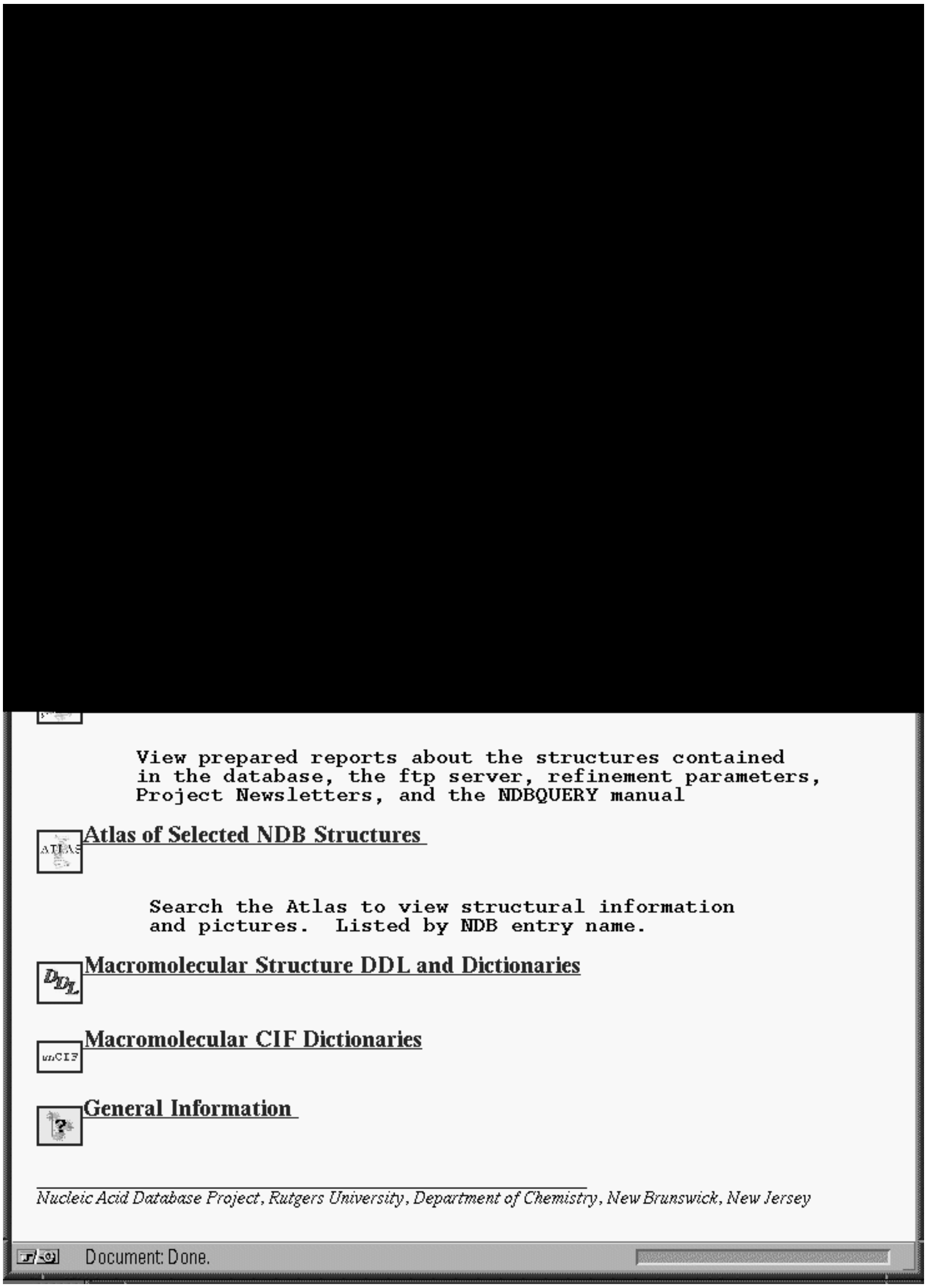




BDL001 over expected ranges for A-DNA.



BDL001 over expected ranges for B-DNA.



View prepared reports about the structures contained in the database, the ftp server, refinement parameters, Project Newsletters, and the NDBQUERY manual

Atlas of Selected NDB Structures

Search the Atlas to view structural information and pictures. Listed by NDB entry name.

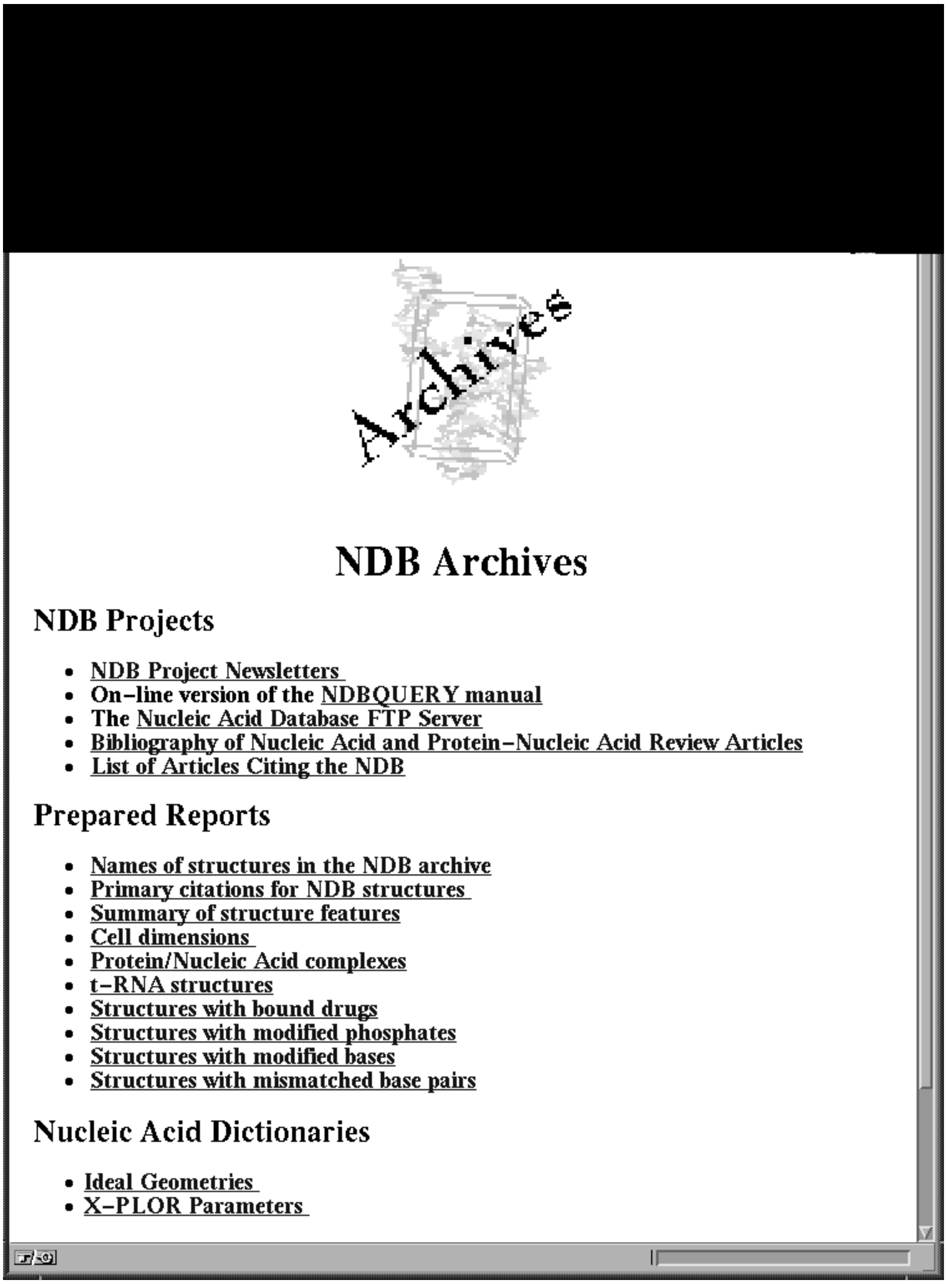
Macromolecular Structure DDL and Dictionaries

Macromolecular CIF Dictionaries

General Information

Nucleic Acid Database Project, Rutgers University, Department of Chemistry, New Brunswick, New Jersey

Fig. 7. The NDB Homepage (available at http://ndbserver.rutgers.edu and is mirrored at the European Bioinformatics Institute at http://www.ebi.ac.uk/NDB/).



[redacted]

NDB Archives

NDB Projects

- NDB Project Newsletters
- On–line version of the NDBQUERY manual
- The Nucleic Acid Database FTP Server
- Bibliography of Nucleic Acid and Protein–Nucleic Acid Review Articles
- List of Articles Citing the NDB

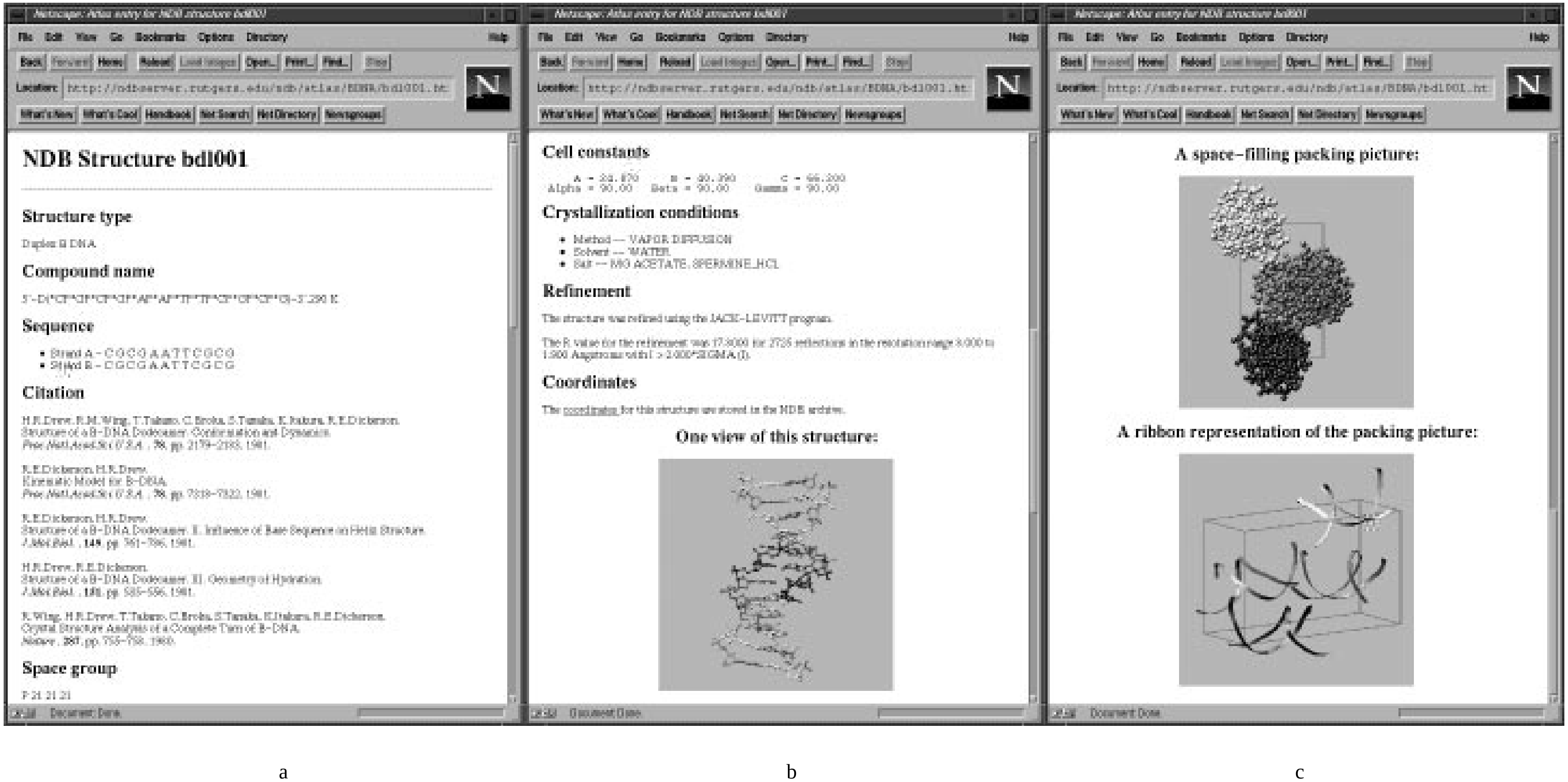
Prepared Reports

- Names of structures in the NDB archive
- Primary citations for NDB structures
- Summary of structure features
- Cell dimensions
- Protein/Nucleic Acid complexes
- t–RNA structures
- Structures with bound drugs
- Structures with modified phosphates
- Structures with modified bases
- Structures with mismatched base pairs

Nucleic Acid Dictionaries

- Ideal Geometries
- X–PLOR Parameters

Fig. 8. The NDB Archives Page.


Netscape: Atlas entry for NDB structure bdl001
File Edit View Go Bookmarks Options Directory Help
Back Forward Home Reload Load Images Open... Print... Find... Stop
Location: http://ndbserver.rutgers.edu/ndb/atlas/BDNA/bdl001.ht
What's New What's Cool Handbook Net Search Net Directory Newsgroups
Document: Done
a
b
c

NDB Structure bdl001

Structure type

Duplex B DNA

Compound name

5'-D(*CP*GP*CP*GP*AP*AP*TP*TP*CP*GP*CP*G)-3',290 K

Sequence

- Strand A - C G C G A A T T C G C G
- Strand B - C G C G A A T T C G C G

Citation

H.R.Drew, R.M.Wing, T.Takano, C.Broka, S.Tanaka, K.Itakura, R.E.Dickerson.
Structure of a B-DNA Dodecamer: Conformation and Dynamics.
Proc.Natl.Acad.Sci.U.S.A., **78**, pp. 2179–2183, 1981.

R.E.Dickerson, H.R.Drew.
Kinematic Model for B-DNA.
Proc.Natl.Acad.Sci.U.S.A., **78**, pp. 7318–7322, 1981.

R.E.Dickerson, H.R.Drew.
Structure of a B-DNA Dodecamer. II. Influence of Base Sequence on Helix Structure.
J.Mol.Biol., **149**, pp. 761–786, 1981.

H.R.Drew, R.E.Dickerson.
Structure of a B-DNA Dodecamer. III. Geometry of Hydration.
J.Mol.Biol., **151**, pp. 535–556, 1981.

R.Wing, H.R.Drew, T.Takano, C.Broka, S.Tanaka, K.Itakura, R.E.Dickerson.
Crystal Structure Analysis of a Complete Turn of B-DNA.
Nature, **287**, pp. 755–758, 1980.

Space group

P 21 21 21

Cell constants

```
    A = 24.870      B = 40.390       C = 66.200
Alpha = 90.00    Beta = 90.00    Gamma = 90.00
```

Crystallization conditions

- Method — VAPOR DIFFUSION
- Solvent — WATER
- Salt — MG ACETATE, SPERMINE_HCL

Refinement

The structure was refined using the JACK-LEVITT program.

The R value for the refinement was 17.8000 for 2725 reflections in the resolution range 8.000 to 1.900 Angstroms with I > 2.000*SIGMA(I).

Coordinates

The coordinates for this structure are stored in the NDB archive.

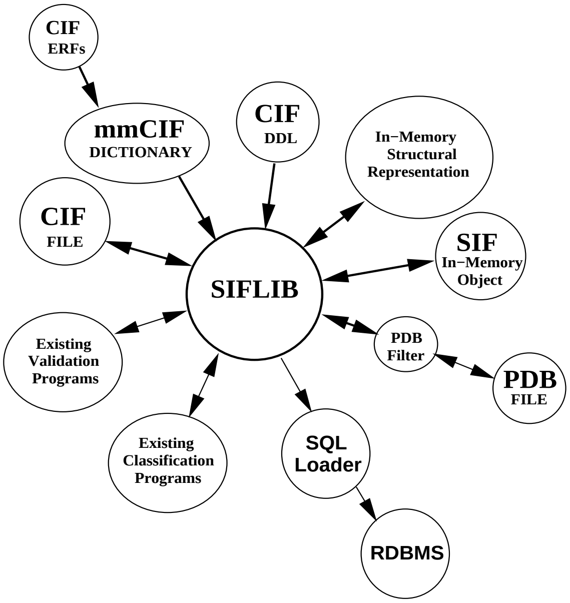
One view of this structure:

A space-filling packing picture:

A ribbon representation of the packing picture:

Fig. 9. An Atlas entry for the first B-DNA crystal structure, BDL001 [13]. (a) The top of the entry page shows the structure type, compound name, sequence, citation, and space group. (b) Also included in the atlas entry are cell constants, crystallization conditions, refinement, and a link to the coordinate file for the structure. A ball and stick representation of the structure is color coded by sequence, with thymine in blue, adenine in red, cytosine in yellow, and guanine in green. (c) The space filling and ribbon representations of the unit cell are color coded in terms of the symmetry related molecules.

various subcategories of structures, including DNA, RNA, nucleic acid-protein complexes, and nucleic acid-drug complexes. Nucleic Acid Dictionaries are included in the NDB Archives, and feature X-PLOR parameters and ideal geometries for DNA/RNA bases



dictionaries; reading and writing individual CIF data items; data integrity checking of CIF data items; and navigation through the CIF schema.

As the first version of the mmCIF dictionary nears completion, the NDB is converting its data processing system based on the mmCIF local dictionary to a system which is based on the data representation in the mmCIF dictionary. The core of this conversion is the integration of SIFLIB into the NDB data processing scheme as shown in Fig. 10. The key feature of this new data processing scheme is that it takes full advantage of the data description provided by the mmCIF dictionary which now contains all of the information necessary to perform detailed integrity checks for individual data items as well as for the relationships between data items.

7.2 Validation

As a result of the surveys of both the NDB and CSD databases, dictionaries of standard covalent geometries and observed ranges of other structural features have been compiled.

These dictionaries provide the foundation for the continued development of structural validation tools that will be used as benchmarks to evaluate each structure submitted to the NDB.

mmCIF provides a mechanism for standardizing the encoding of structural standards and other lengthy tabulations reference data in External Reference Files (ERFs). Information stored in ERFs can be accessed using the same software (SIFLIB) as other CIF data. We plan to integrate structural ERFs automatically into the NDB data processing scheme (Fig. 10).

7.3 Information Retrieval

The recently developed WWW interface to the NDB database provides the structure selection features of the more robust menu-driven interface, NDBquery. An enhanced version of the WWW interface that will provide both structure selection as well as report generation has recently been released.

The WWW interface is shown schematically in Fig. 11. The figure highlights the underlying use of a CIF dictionary to describe the database schema for the WWW interface.

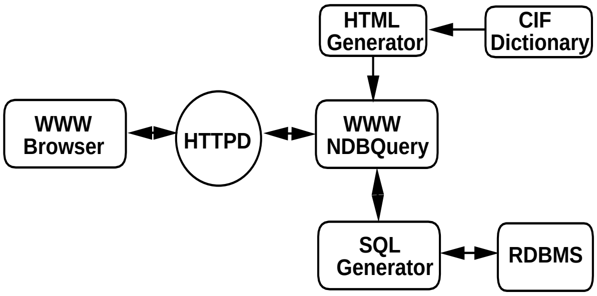


Fig. 11. Schematic view of the NDB WWW forms based interface. The WWW version of NDBquery is called by the WWW server and provides the server with a description of the contents of the NDB database, which is presented as a set of menu selections. The WWW interface also manages the construction of SQL queries and all communication with the NDB database.

Acknowledgments

This work is supported by the NSF (BIR 9510703). The collaboration with Wilma Olson and David Beveridge on nucleic acid analysis projects using the NDB is gratefully acknowledged.

8. References

[1] H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider, The Nucleic Acid Database—A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids, Biophys. J. **63** (3), 751-759 (1992).

[2] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures, J. Mol. Biol. **112,** 535-542 (1977).

[3] S. Macskassay, Ndbview. A Specialized 3-Dimensional Display Program for Crystallographic Structures of Nucleic Acids., Rutgers University, New Brunswick (1991).

[4] F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hunnelink, O. Kennard, W. D. S. Motherwell, J. R. Rogers, and D. G. Watson, The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information, Acta Cryst. **B35,** 2331-2339 (1979).

[5] L. Clowney, S. C. Jain, A. R. Srinivasan, J. Westbrook, W. K. Olson, and H. M. Berman, Geometric Parameters In Nucleic Acids: Nitrogenous Bases, J. Am. Chem. Soc. **118**, 519-529 (1996).

[6] A. Gelbin, B. Schneider, L. Clowney, S.-H. Hsieh, W. K. Olson, and H. M. Berman, Geometric Parameters In Nucleic Acids: Sugar and Phosphate Constituents, J. Am. Chem. Soc. **118**, 509-518 (1996).

[7] G. Parkinson, J. Vojtechovsky, L. Clowney, A. T. Brünger, and H. M. Berman, New Parameters for the Refinement of Nucleic Acid Containing Structures, Acta Cryst. D, 52, 57-64 (1996).

[8] B. Schneider, S. Neidle, S.-H. Hsieh, and H. M. Berman, A Comprehensive Analysis of the Sugar-Phosphate Backbone in Helical DNA Crystal Structures, J. Mol. Biol submitted, (1996)..

[9] J. D. Westbrook and S. S. Hall, A Dictionary Description Language for Structure Macromolecular, Rutgers University (1994).

[10] S. R. Hall, F. H. Allen, and I. D. Brown, A New Standard Archive File for Crystallography, Acta Crystallogr. **A47,** 655-685 (1991).

[11] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson, Structure of a B-DNA Dodecamer: Conformation and Dynamics, Proc. Natl. Acad. Sci. U.S. **78** (4), 2179-2183 (1981).

[12] K. Grzeskowiak, K. Yanagi, G. G. Privé and R. E. Dickerson, The Structure of B-Helical C-G-A-T-C-G-A-T-C-G and Comparison with C-C-A-A-C-G-T-T-G-G. The Effects of Base Pair Reversals., J. Biol. Chem. **266,** 8861-8883 (1991).

[13] A. R. Srinivasan and W. K. Olson, Viewing Stero Drawings, J. Chem. Ed. **66,** 664-665 (1989).

About the authors: *Helen M. Berman is the Head of the NDB Project. John Westbrook, the database designer; Anke Gelbin, the data coordinator; Les Clowney and Shu-Hsin Hsieh, programmers for the project; and*